

The invention in which an exclusive right is claimed is defined by the following:

1. A method of distributing a processing load in a cluster having a plurality of resources, comprising the steps of:

- (a) designating a first resource as an intake for a first session;
- (b) directing a plurality of new client requests for service by the cluster to the intake to form a first group of clients, wherein each client in the first group continues to receive services only from the first resource for as long as those services are provided;
- (c) determining that a second resource be designated as a new intake, to balance the processing load among the plurality of resources;
- (d) designating the second resource as the new intake; and
- (e) directing successive new client requests for services by the cluster to the new intake to form a second group of clients, wherein each client in the second group continues to receive services only from the second resource for as long as those services are provided.

2. The method of Claim 1, wherein the step of designating a first resource as an intake comprises the steps of:

- (a) assigning a unique identifier to each resource in the cluster; and
- (b) selecting the resource that will be designated as a function of its identifier.

3. The method of Claim 1, wherein the step of designating a first resource as an intake comprises the steps of:

- (a) calculating a rating value for each resource in the cluster; and
- (b) selecting the resource that will be designated as a function of the rating value.

4. The method of Claim 1, wherein the step of designating a first resource as an intake comprises the steps of:

- (a) calculating a time-out; and
- (b) selecting the resource that will be designated as a function of the time-out.

5. The method of Claim 1, wherein the step of directing the plurality of new client requests for service to the intake to form the first group of clients comprises the steps of:

- (a) receiving a request for service from a new client, wherein the request is received by a resource other than the intake; and
- (b) directing the client to the intake.

6. The method of Claim 1, wherein the step of directing the plurality of new client requests for service to the intake to form a first group of clients comprises the steps of:

- (a) receiving a request for service from a new client, wherein the request is received by a resource other than the intake; and
- (b) transferring the request for service by the new client to the intake.

7. The method of Claim 1, further comprising the steps of:

- (a) detecting a termination in a service being provided to a client by one of the plurality of resources;
- (b) determining that the client is requesting a service from the cluster; and
- (c) directing the client to a current intake for the service requested by the client.

8. The method of Claim 1, wherein the step of determining that the second resource be designated comprises the steps of:

- (a) calculating a load value of the first resource;
- (b) comparing the load value to a threshold value; and
- (c) designating the second resource as the new intake, if the load value exceeds the threshold value.

9. The method of Claim 1, wherein the step of designating the second resource as the new intake comprises the steps of:

- (a) calculating a rating value for each resource in the cluster; and
- (b) selecting the resource that will be designated as a function of the rating value.

10. The method of Claim 1, further comprising the step of periodically exchanging status messages between the plurality of resources, wherein the step of determining that the second resource be designated occurs if a status message has not been received from the intake within a predetermined period of time.

11. The method of Claim 10, wherein the step of designating the second resource as the new intake comprises the step of the second resource assuming the designation as the new intake after the second resource fails to receive the status message from the first resource within the predetermined period of time, said status message identifying the first resource as the intake.

12. The method of Claim 1, wherein the step of designating the second resource as the new intake comprises the steps of:

- (a) providing an intake message from the first resource to the plurality of resources in the cluster identifying the second resource as the intake;
- (b) receiving the intake message at the plurality of resources in the cluster, including the second resource; and
- (c) updating a list at each of the plurality of resources in the cluster, said list indicating that the second resource has been designated as the new intake.

13. The method of Claim 12, further comprising the step of providing a message from the second resource to the plurality of resources in the cluster identifying the second resource as the new intake to confirm that the second resource has accepted its designation as the new intake and to ensure that the plurality of resources are aware of the new intake.

14. The method of Claim 1, wherein the cluster comprises a plurality of nodes on which the plurality of resources are implemented, and wherein the step of designating the second resource as the new intake comprises the steps of:

- (a) determining that the second resource and first resource reside on a common node;
- (b) updating a list stored on the common node, said list indicating that the second resource is designated as the intake; and
- (c) providing a message from the second resource designating the second resource as the intake.

15. The method of Claim 1, further comprising the step of the first resource providing a data message to the plurality of resources in the cluster, said data message including an identification of the first resource and a load value of the first resource.

16. The method of Claim 1, further comprising the following steps that are carried out by a client:

- (a) storing a network address for one resource in the cluster;
- (b) automatically attempting to connect to said one resource at the network address;
- (c) receiving from the cluster a network address for the intake for a service requested by the client; and
- (d) automatically attempting to connect to the network address for the intake.

17. A machine readable medium having machine-executable instructions for performing the steps of Claim 1.

18. A machine readable medium having machine-executable instructions for performing the steps of Claim 16.

19. A system for distributing a processing load in a cluster, comprising:

- (a) at least one processor for implementing the cluster, said at least one processor comprising a plurality of resources that provide services to a plurality of clients;
- (b) an interface for coupling said at least one processor to the plurality of clients;
- (c) a memory in which a plurality of machine instructions are stored; and
- (d) said machine instructions, when executed by said at least one processor implementing:
 - (i) a first resource operatively connected to the plurality of clients, said first resource being designated as an intake that accepts requests from new clients for a service, and in response thereto, forming a first group of clients that continue to receive services only from the first resource for as long as those services are provided;
 - (ii) said first resource determining to designate a second resource from among the plurality of resources as a new intake, the second resource being connected in communication with the first resource; and
 - (iii) designating the second resource as the new intake to accept new client requests for service, forming a second group of clients that continue to receive services from the second resource for as long as those services are provided.

20. The system of Claim 19, wherein the machine instructions further cause a new client request for service to be directed to a resource currently designated as the intake.

21. The system of Claim 19, wherein the machine instructions are executed in a plurality of instances by a plurality of processors.

22. The system of Claim 19, wherein a first instance of the machine instructions for load balancing are executed to manage the first resource and a second instance of the machine instructions for load balancing are executed to manage the second resource, said machine instructions causing said first instance to communicate with said second instance, and wherein said first instance of the machine instructions cause the first resource to transfer the intake designation to the second resource.

23. The system of Claim 19, further comprising a client device having a client processor and a client memory in which are stored:

- (a) machine instructions; and
- (b) a list that includes at least one network address corresponding to at least one resource in the cluster, said machine instructions stored in the client memory causing the client processor to:
 - (i) automatically attempt to connect to said at least one resource using the network address corresponding thereto;
 - (ii) receive from the cluster an intake network address corresponding to a resource designated as the intake for said at least one service; and
 - (iii) automatically attempt to connect to the intake network address.

24. A method of distributing a processing load among a cluster of nodes, each node providing at least one of a plurality of different types of services, comprising the steps of:

- (a) designating a first instance of a first type of service on a first node as an intake;
- (b) directing new client requests for said first type of service to the intake to form a first group of clients, wherein each client in the first group continues to receive services only from the first instance on the first node for as long as those services are provided;
- (c) determining that a second instance of the first type of service be designated as a new intake for the first type of service;

(d) designating the second instance as the new intake for the first type of service; and

(e) directing a plurality of new client requests for the first type of service to the new intake to form a second group of clients, wherein each client in the second group continues to receive services only from the second instance for as long as those services are provided.

25. The method of Claim 24, wherein the step of directing new client requests for said first type of service to the intake to form a first group of clients comprises the steps of:

(a) receiving from a new client a request for said first type of service, wherein the request is received at a node other than the node on which the intake is designated; and

(b) directing the client to the intake.

26. The method of Claim 24, wherein the step of directing a plurality of new client requests for service to the intake to form a first group of clients comprises the steps of:

(a) receiving from a new client a request for said first type of service, wherein the request is received at a node other than the node on which the intake is designated; and

(b) transferring the request for service by the new client to the intake.

27. The method of Claim 24, wherein the step of determining to designate a second instance as the new intake comprises the steps of:

(a) calculating a load value for the first node, said load value being normalized to enable a uniform comparison to corresponding load values for the other nodes of the cluster;

(b) comparing the load value for the first node with a threshold value; and

(c) designating the second instance as the new intake, if the load value exceeds the threshold value.

28. The method of Claim 24, wherein the step of designating the second instance as the new intake for the first type of service comprises the steps of:

(a) calculating a rating value for each resource in the cluster; and

(b) selecting the resource that will be designated as a function of the rating value.

29. The method of Claim 24, further comprising the step of periodically exchanging status messages between the plurality of nodes, wherein the step of determining that the second resource be designated occurs if a status message has not been received from the intake within a predetermined period of time.

30. The method of Claim 29, wherein the step of designating the second instance as the new intake for the first type of service comprises the steps of a second node assuming authority to designate the second instance as the new intake; and automatically selecting the second instance as the new intake from a plurality of instances of the first type of service on the second node after the second node fails to receive the status message from the first instance within a predetermined period of time, said status message identifying the first service instance as the intake.

31. The method of Claim 24, wherein the step of designating the second instance as the new intake for the first type of service comprises the steps of:

- (a) providing an intake message from the first node holding the first instance to the nodes in the cluster identifying the second instance as the intake;
- (b) receiving the intake message at the nodes in the cluster, including a second node on which the second instance is executing; and
- (c) updating a list at the nodes in the cluster, said list indicating that the second instance has been designated as the new intake.

32. The method of Claim 31, further comprising the step of providing a message from the second node to the nodes in the cluster, said message identifying the second instance as the new intake to confirm that the second instance has accepted its designation as the new intake and to ensure that the plurality of nodes are aware of the new intake.

33. The method of Claim 24, wherein the step of designating the second instance as the new intake comprises the steps of:

- (a) determining that the second instance and first instance reside on a common node;
- (b) updating a list stored on the common node, said list indicating that the second instance is designated as the new intake; and
- (c) providing a message from the common node to the nodes in the cluster, said message identifying the second instance as the new intake.

34. The method of Claim 24, further comprising the step of providing a data message from the first node to the plurality of nodes in the cluster, said data message including an identification of the first instance and a load value of the first node.

35. The method of Claim 24, further comprising the step of sending a service message from the first instance to a control process executing on the first node, said service message including a unique identification of the first instance and operational status parameters of the first instance that the control process uses to calculate a rating value for the first instance and a load value for the node that are used to determine a future intake designation.

36. A system for distributing a processing load in a cluster of resources, comprising:

(a) means for enabling communication between the resources comprising the cluster;

(b) means for enabling communication between the resources comprising the cluster and a plurality of clients requesting services from said resources;

(c) means for designating a first resource as an intake so that the first resource accepts requests from new clients for a service, and in response thereto, forms a first group of clients that continue to receive services only from the first resource for as long as those services are provided;

(d) means for determining to designate a second resource as a new intake; and

(e) means for designating a second resource as the new intake so that the second resource begins to accept requests from new clients for the service, and in response thereto, forms a second group of clients that continue to receive services only from the second resource for as long as those services are provided.